

**PDF-Archive  
Draft Meeting Minutes  
May 1, 2003**

1. Introductions

Mr. Stephen Levenson and Mr. Dana Stone called the meeting to order at 9:05 a.m.  
The meeting participants included:

Mrs. M. Abbott	NPES
Mr. S. Abrams	Havard University Libraries
Ms. Christina Bontempo	NPES
Mr. Robert Breslawski	Eastman Kodak
Ms. Jane Cohen	DTIC
Ms. Marian Ellis	Merck & Company
Ms. Betsy Fanning	AIIM International
Mr. Scott Foshee	Adobe
Mr. Mark Gavin	Appligent, Inc.
Mr. Kenneth Haller	Tessada & Associates
Mr. John Janick	Merck & Company
Mr. Stephen Levenson	Admin. Offices U.S. Courts
Mr. William LeFurgy	Library of Congress
Mr. Joe McConnell	ProQuest Co.
Mr. David McDowell	NPES
Mr. Basil Manns	Library of Congress
Ms. Pamela Mason	NARA
Mr. Jeffrey Reed	NARA
Mr. Leonard Rosenthol	PDF Sages Inc.
Mr. Rick Seeler	Adobe
Mr. Paul Showalter	Internal Revenue Service
Mr. Joe Speetjens	BAE Systems
Mr. Dana Stone	Merck & Company
Ms. Susan Sullivan	NARA
Mr. Arnold Thielen	MIXNET/eFile.com
Mr. Doug Vander Wilt	
Ms. Melonie Warfel	Adobe

2. Approval of May 1, 2003 Agenda

The May 1, 2003 meeting agenda (PDF-A 2003-004) was approved with the addition of discussion points for intellectual property and security under agenda item 4. A discussion on the introduction and scope was requested after the approval of the minutes and prior to the working group reports. The chairperson for the Forms working group was changed from Ms. M. Warfel to Mr. M Gavin. (McDowell/Warfel)

### 3. Approval of February 27-28, 2003 Meeting Minutes

The draft meeting minutes for the February 27-28, 2003 (PDF-A 2003-003) meeting were approved as amended. (McDowell/Martin) On page 3, the word "recommended" was changed to "discussed". (Cohen/Sullivan) NARA proposed changes to the minutes that were not accepted.

The comments submitted by NARA and KODAK regarding the Introduction and Scope were discussed. It was agreed that "creation" would be changed to "rendering" throughout the scope. It was also noted that the scope statement is intended to be a broad statement. Some changes were made to the scope based on the recommendations from NARA and KODAK. In the interest of time, a small working group, chaired by Susan Sullivan, was formed to review the scope, introduction and title. The concepts of 'long term' and 'preservation' are core to the document, restrictions and decisions made on that basis.

### 4. Reports and Review of Draft

As each working group reported on the progress of the work the committee discussed the related section of the document and modified the document as appropriate.

Mr. Abrams reported on the conformance section that is based on discussion from the previous meeting. This section specifies both a minimal conformance profile and a full conformance profile. It was suggested that the standard use the term "level" rather than "profile" since profile is used in various ways in various industries (e.g., color management profiles). In addition, there was concern that every item would need to specify to which conformance level it applied. It was suggested that the conformance section needs to first identify what is needed for the full conformance level, but can then list the items that can be omitted in a minimal conformance level. This distinction will need to be made clear especially as it applies to file format.

The committee discussed using the term "consumers" rather than "readers," since readers are often thought of as display tools. A consumer may be a validation tool or some other device that is not intended to render the document in any way. It was agreed to add the definition of reader as defined in ISO 15930-3 to read: "software application that is able to read and appropriately process files". A revised draft of the Conformance section will be developed.

Mr. Gavin reported on the file format working group. There are many creation tools that build and/or manipulate PDF files; however, some do not do a good job. However, there are some PDF tools that will read a PDF file that does not actually conform to the PDF Reference Manual. Therefore, the proposed wording was developed with a goal of ensuring ease of reading in the future, using Adobe Acrobat as a model. It was agreed that we must be specific about what requirements are restricting what the PDF Reference Manual already allows. For example, it was agreed to require that %PDF be the first four characters of the header, and these characters shall not be preceded by any other characters or spacing.

In section 5.1, it was recommended that "Cos" be removed throughout.

In section 5.2, there are three methods of signaling a line ending: single carriage return, single line feed, or a combination of the two. It is not uncommon to have a file that contains a variety of line endings. These often differ according to the operating system used. For the sake of consistency, it was proposed that the two-character combination of "carriage-return, line feed" be required to indicate the end of a line. It was also agreed that the requirements be the first paragraph, with the rest being a note.

In section 5.3, an explanatory note was added to read: These requirements deal with white space as far as they relate to the file format structure and not the content of the document.

In section 5.4, it was questioned whether the standard should specify what binary data should be used.

In section 5.5, it was suggested the InfoRange Array, MetaDataRange Array and MetaData Dictionary be removed to allow search engines to more easily find metadata information within a PDF/A file. It was also agreed to remove 5.1 on range arrays.

It was noted that a PDF/A file may not be encrypted; therefore, the keyword Encrypt shall not be included in a PDF/A document trailer. It was suggested this be reiterated in an encryption section within the standard. It was also questioned whether there should be a separate section addressing security, since prohibiting encryption addresses most of the security issues. It was felt the only other issue would be a digital signature. It was suggested that if there are other security issues, including digital rights, they can be handled in an informative annex that can point to the relevant portions of the standard.

It was felt the contents of 12.6 (Compression of images and text) and 12.7 (Encryption and digital signatures) can be handled in other sections of the document and further discussed in an informative annex relating to security.

In section 5.7, the committee discussed whether or not to allow updated PDFs. The committee discussed the possibility of allowing incremental updates where there could be multiple versions of the metadata. There will be a need to identify the intended metadata without first parsing the file. If the standard allows updated documents, one issue is that any reader that is only looking at metadata (for example) would need to be more complex. It was suggested the standard allow incremental updates. When updating a file, one must also update the XMP. Therefore it was questioned whether or not to prohibit the Document Information Dictionary. If both XMP and the Info Dictionary are allowed, there must be a way to ensure they are in sync. Therefore, it was agreed to prohibit the Document Information Dictionary. It was noted that when older files are written as PDF/A files, the writer will convert the information in the Info Dictionary to XMP and delete the Info Dictionary.

It was agreed to remove "Changes array" from Table 2.

It was agreed to remove 5.8 (Change dictionary)

In section 5.9, the heading for the section was changed to "Extraneous binary data." The first sentence was changed to read, "No data shall exist following the final %%EOF at the end of the line."

In section 5.10, "Cos object" was changed to *literal string*.

In section 5.11, "Cos name" was changed to "Name objects". Wording was added to state that empty names shall not be permitted.

In section 5.12, "Cos stream" was changed to *stream object*. It was agreed that the means of specifying the stream length needs to be specified.

In section 5.13, the first paragraph was converted to an explanatory note.

In section 5.14 (Balanced page trees), it was also agreed that the section on balanced page trees should be handled as part of an informative annex on ways to write better PDF/A files or in application notes.

In section 5.15 (Linearization), it was agreed to state that linearized PDF files do not have to be recognized by a conforming reader, and the first sentence was made an informative note.

In section 5.16 (Binary file format), it was agreed to delete this section.

In section 5.17 (Filters), the first paragraph was deleted. Also deleted the first sentence of the 3<sup>rd</sup> paragraph.

In section 5.18 (Character encoding), this section was deleted.

In section 5.19 (Streams), was changed to "stream object."

A new section (5.2 Document catalog) was added to identify a version key. Wording for this section will be developed.

It was agreed to table the Graphics and Transparency sections in light of Mr. Hughes absence. However, the committee briefly discussed section 6, Graphics and made the following modifications and decisions.

Section 6.1 (General), transparency will be moved.

In section 6.10 (Content streams), the sentence was broken into two parts, with the first sentence disallowing any operators in the content stream that are not documented in the PDF Reference Manual, and the second sentence recommending that the content stream not contain other data not documented in the PDF Reference Manual.

Mr. Abrams reported on the Fonts working group making the following changes to the document.

In section 7.4 (Embedded font programs), the first sentence of Note 1 was moved into the normative wording. It was also agreed to allow hidden text. The first paragraph was edited to read "All Type0, Type 1, and TrueType fonts referenced for render within a PDF/A file shall be embedded with that file, including the 14 standard Type 1 fonts, if they are used.

"NOTE: An example of a font referenced but not rendered is text mode 3 (invisible)."

The committee discussed what happens when a font is used that is not "legally" embedded. It was noted that because of the requirements of 7.4, such a file would not be a valid PDF/X-A file.

In section 7.7 (Font metrics), was changed to read "For all embedded fonts, a compliant ..."

In section 7.9 (Unicode character maps), it was noted that this applies only for the full conformance level.

Mr. LeFurgy reported on the work of the XML/Metadata working group. The following changes were made to the document based on the committee discussion.

In section 11.2 (Properties), the committee debated whether or not to remove the last sentence, which reads, "The metadata stream shall be visible as plain text to non-PDF/A aware tools and so shall be both unfiltered and unencrypted." Following the discussion, the committee voted to retain the sentence, with a minor edit to the end of the sentence.

In section 11.3 (Normalization), it was reported that Adobe noted that some of these normalizations that were taken out of the XMP specification are incorrect, and Adobe is supposed to provide feedback on this issue.

In section 11.6 (File provenance information), it was noted that this section provides for, but does not require the use of, a metadata audit trail. However, some questioned the need for a secondary audit trail, and there was concern this may add unnecessary complexity. This secondary audit provides repeatable versions of metadata that reflect the various stages the document has passed through. For clarity, it was explained that the first paragraph does not really describe a true audit trail, but is a history of the steps taken to create the file. Revised wording to clarify this section will be proposed for review at the next meeting. It was agreed to reference ISO 8912 for the format for the timestamp. It was also noted that the standard should address how to store this information, if you have it.

In section 11.7 (Use of non-XMP metadata), it was agreed to disallow non-XMP metadata. It was also agreed to move this section to the annex on how to create better PDF/A files.

In section 11.8 (Extension schemas), it was reported that Adobe expressed interest in supporting the RDF Schema Specification and will look into this further. This section recommends that for each PDF/A file you write, the XML schema used should also be embedded in the file. This will facilitate validation of metadata.

In section 11.9 (Validation), it requires that the file be validated for conformance with the XML/RDF syntax, which is currently possible. It also says one should validate for proper values and data types, which is not yet supported by PDF and is under discussion with Adobe. It was suggested a set of questions that are being raised be captured and taken back to experts who can address them.

It was agreed to delete section 11.11 (Character property metadata).

It was agreed to delete section 11.12 (Natural language private use identifier metadata).

Mr. Gavin reported on the forms working group conference call. A digital signature is built on a form field, so there must be a portion of a form field (which will not necessarily be visible) in order to have a digital signature. The encryption and digital signature issues in 12.7 will be addressed in other sections, so 12.7 was removed. The same is true of 12.6.

For every form there are two parts, the value of the field and the appearance of the field. The field data (value) is not used for rendering, but only for data retrieval. The appearance field is what is used for rendering. There is a need to capture this concept more clearly in the text. An example is the case in which a form has drop down options. Should the options remain "live", or should the archive be the document showing only the choices selected, with the capability to go back to the metadata to view what options were offered from which to choose.

There needs to be continued discussion on forms and on digital signatures. One of the issues with digital signatures is the ability to preserve digital signatures that have been applied to the document prior to conversion to PDF/A. Currently, PDF does not provide for this, but it is hoped that a future version will facilitate this. The other issue is that in some cases it is not desirable to have the archived document "live" with the ability to add digital signatures after it has been prepared for archival. There may also be a need to ensure that there is a way to know the document has not been changed after it was "signed."

The format issue is that there can be multiple workflows, and the format needs to enable these workflows. The committee should discuss the workflow of when and where a PDF/A document would be created. A suggestion was made to make PDF-A documents read-only but no consensus was reached on that issue.

## 6. Wrap-up and Next Steps

The next meetings of the committee will be:

July 10-11, 2003	Washington, DC
September 4-5, 2003	Washington, DC
October 20-24, 2003	New Orleans, LA (ISO Meeting – only one day)
November 20-21, 2003	Washington, DC

## 7. Adjournment

Prior to adjourning, Mr. Levenson acknowledged and thanked Mr. Dana Stone for his contributions to the committee as he is stepping down as the co-chair of the committee but will be following the progress of the committee. This change is due to new job responsibilities assigned to Mr. Stone. The committee thanked Mr. Stone.

The meeting was adjourned at 5 p.m.